

4 *Some case studies on linguistic variation and their implications*

CAROL GENETTI

Introduction

One of the joys of reading the work of James A. Matisoff is his own clear delight in the messiness and infinite variation in language. Rather than trying to whittle a language down in order to achieve a single monolithic description, he fleshes it out, deftly exploring the subtlety and variation, and revealing the power of creativity of native speakers. In my own experience of conducting fieldwork on languages of the Himalayas, I have found myself repeatedly confronted by linguistic variation of a number of types and with a number of motivations. In my earlier days, I was eager to attribute such variation exclusively to differences of dialect (geographically or socially defined) or register. But eventually I had to admit that I was confronting variation that was neither, but was variation at the level of the idiolect. This confounded my ability to create the monolithic description that I thought grammars were supposed to be, and eventually, under the influence of Matisoff's work, I learned to work with the variation, indeed to give it a central role in my understanding of language.

For this volume, which honors Jim Matisoff and his tremendous accomplishments, I have decided to bring together several case studies of variation taken from my own work. The first study is on verb agreement in Nepāli, and presents a classic case of variation based on register. It shows that variation can persist across generations, and that the register-based system is the result of competing pressures on the system. The second study examines differences in syntactic constructions between two speakers of Kāthmāndu Newar who otherwise show very little difference in their speech patterns. While it is certainly possible that one could attribute these differences to dialect, probably socially defined, it is also possible that the differences are idiolectal. The third study examines

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 53–63.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

idiolectal variation in more detail, comparing differences in the speech of first cousins raised as young children in the same household. Idiolectal differences in their phonological systems occur in environments where the differences carry no functional load. I also discuss differences in their use of two paradigmatically related suffixes, and suggest that the variation is the result of a decayed, relatively unmotivated distribution.

Throughout the paper, I will be referring to an expectation that is widely held in linguistics: the expectation of inter-speaker consistency. It is generally expected that when two or more speakers speak the same dialect, their phonological and grammatical systems will be the same. Thus one assumes that working with one speaker as an informant will produce the same result as working with another speaker, although there may be differences between them in their skill as informants, in the size of their vocabulary, etc. When variation is encountered, one can attribute it to dialect or register, and still happily feel that inter-speaker consistency is maintained. Part of the motivation behind the expectation of inter-speaker consistency comes from our tradition as grammarians to come up with a single description of each component of a language, and to present these descriptions as invariable, and as representing 'the way' a language works. It also derives in part from our tradition of conducting elicitation with a single, primary consultant, which naturally precludes any evidence of inter-speaker variation. And of course it is also due to the fact that there is ample evidence that social factors do indeed correlate with variation.

While the first study discussed below comfortably allows us to maintain the expectation of inter-speaker consistency, as variation is attributed to register differences, the second study raises doubts, and the third study, on idiolectal variation, runs counter to the expectation. One of the goals of this paper, then, is to make this expectation explicit, which is often implicit and assumed, and to demonstrate that it does not necessarily hold true. The second goal is to illustrate the richness of analysis which comes from the exploration of variation in detail. The historical sources and synchronic implications of linguistic variation, even at the level of the idiolect, are themselves a fascinating and enriching field of study.

1 Verb agreement in Nepāli: register-based variation persisting across generations

The first case study concerns variation in verb agreement in Nepāli, an Indo-Aryan language which is the national language of Nepal. A full discussion can be found in Genetti (1999); here I will only summarise that study and briefly discuss the results.

Nepāli is a language with many varieties and a rich literature. There is a strong prescriptive tradition, and written Nepāli, which is taught to children in schools, is taken to be the 'correct' form of the language. In this prescriptive variety, the verb agrees with the subject in person, number, gender, and honorific status, all of which combine to create a

complex verbal paradigm (Genetti 1999:543–544). However, in casual spoken Nepāli, native speakers will frequently simplify the agreement, in particular the gender and number inflection. This variation is quite noticeable. In Genetti (1999), I presented clear evidence that amount and type of agreement is dependent on genre. In written Nepali there are very high percentages of agreement in gender and number (97.3 per cent, combined), in conversational Nepāli, percentages of agreement are quite low (9.3 per cent), and in spoken Nepāli narratives, the verb agreed in gender and number about half the time (51.9 per cent). Interestingly, the most inter-speaker variation found within genre was attested in the narrative data, the register intermediate in its formality. Here the percentages of different speakers varied significantly, but this variation could be attributed to different interpretations of formality that different speakers brought to the task of telling a story to a linguist with a tape recorder. Thus, in this case one can still maintain the expectation of inter-speaker consistency. Variation is attributed to register, and there is the expectation that if the speakers were consistent in their interpretation of level of formality, we would also find greater consistencies in their percentages of verb agreement.

One cannot help but wonder about the motivations for this register-based variation in the grammatical system. Where does it come from? And, if Nepāli speakers use verb agreement so sparingly in conversation, could it be that it is being lost, and that the current stage is a step away from the elaborate finite paradigm? There are strong internal and external pressures that would favor such a change. Internal pressures include the marked status of the feminine and plural categories in the structure of the finite and non-finite paradigms, as well as the low frequency with which these categories appear. The external pressure is substratum interference. Large numbers of speakers of Tibeto-Burman languages learn Nepāli as a second language, and use it as a lingua franca. Gender is not marked on the verb in any of the Tibeto-Burman languages of Nepal, and number is marked in only some of them. A large number of these speakers will learn the verb agreement system imperfectly, and simplify the paradigm, especially the gender and number agreement. This simplification is then an additional motivation for the simplification of the system by native speakers of Nepāli, who are thus frequently exposed to verb forms where the verb does not fully conjugate. This type of substratum interference is not the classic case of speakers shifting from one language to another (see, for example, Thomason & Kaufman 1988), but comes instead from the continued imperfect bilingualism of second-language learners.

While one might be inclined to think that the agreement system in the process of simplification in modern Nepāli, and that this variation represents historical change in progress, there is evidence that this is not a recent phenomenon. The earliest set of spoken Nepāli texts was published by Sir Ralph Lily Turner just after World War I (1921, 1922). They are personal narratives of Gurkha soldiers and their experiences fighting in the war. Due to the nature of the subject matter, there are no feminine referents, hence no data on gender agreement. However, there is evidence for number agreement: in these narratives

the verb agrees in number with inanimate plural subjects only 20 per cent of the time, while it agrees with animate plural subjects about 75 per cent of the time. We can see that lack of perfect agreement in the verb was a feature of Nepāli almost a century ago.

What we see from this study is that the Nepāli system of register-based variation in agreement has persisted over generations, and most likely will continue to persist into the future. The persistency of the variation is due to competing motivations. There is a strong motivation for a simplification of the system, with both internal and external pressures at work. As a counter-balance, there is the strong prescriptive tradition: the complex paradigm is taught in schools, enforced in published Nepāli writings, used in much of the Nepāli media, and used in formal contexts, such as academic lectures. These competing motivations are the source both of the register-based nature of the system, and the persistence of the variation over time.

2 Variation within the Kāthmāndu Newār community

In my early work on Kāthmāndu Newār, before my first trip to Nepal, I worked with two Newārs who were students at the University of Oregon. In most respects, the phonological and grammatical systems of the two speakers were the same; I only began to discover significant differences between them when exploring syntactic constructions in depth. Both speakers exhibited a difference between a ‘long participle’ and a ‘short participle’ form. The long participle form is used by both speakers for narrative chaining, and is often accompanied by a distinctive intonation contour. The short participle is used for the incorporation of auxiliaries (all of which are versatile verbs) into the clause. Thus, both speakers differentiate the following sentences:¹

- (1) *w-āā* *nay-āā* *con-a*.
 3s-ERG eat-l.PART stay-PST.DISJUNCT
 ‘He ate and stayed.’
- (2) *w-ā* *nay-ā* *con-a*.
 3S-ERG eat-S.PART stay-PST.DISJUNCT
 ‘He was eating.’

For one speaker, Rajendra, the short participle is only used for the incorporation of versatile verbs as auxiliaries. The other speaker, Manoj, however, also uses the short participle in a very restricted construction with motion verbs. In this construction, the V2 of the sequence must be either *wan-e* ‘go’ or *wa-ye* ‘come’, and the motion verb must be the only member of its clause; intervening arguments, locations, or adverbials are

¹ The following abbreviations are used in this paper: ERG ergative; LOC locative; L.PART long participle; NEG negative; NR nominalizer/relativizer; PART participle; PL plural; PST past; S.PART short participle; STAT stative.

prohibited. In this construction the focus is on the first verb, but the speaker insists that the meaning of the motion verb is also distinctly conveyed:

- (3) *jī-ī* *nay-ā* *wayā*
 1S-ERG eat-S.PART come-PST.CONJUNCT
 'I already ate (before I came)'

It is interesting that the distinction between the presence and absence of the third, intermediate, construction correlates with another distinction between these two speakers, that of the interpretation of the scope of negation in clause chains. Rajendra, the speaker who lacks the intermediate construction, is quite free in his interpretation of the scope of negation in clause chains, allowing negation on the final verb to be applied to non-final clauses. Thus compare the non-negated sentence in (4) with the negated sentence in (5):

- (4) *jī-ī* *bārcā* *kurk-āā* *tachyān-ā*
 1S-ERG bowl drop-L.PART break-PST.CONJUNCT
 'I dropped the bowl and broke it.'
- (5) *jī-ī* *bārcā* *kurk-āā* *ma-tachyānā*
 1S-ERG bowl drop-L.PART NEG-break-PST.CONJUNCT
 'I dropped the bowl but didn't break it.' OR 'I broke the bowl without dropping it.'

Not any combination of clauses may have backward spreading of negation in this way, rather the clauses must indicate events that are thematically continuous, and both clauses must have the same subject referent.

In contrast, Manoj does not allow an interpretation of example (5) with the focus of negation on the non-final clause. The only conditions under which Manoj will allow backward scope of negation are the same conditions that hold for his intermediate chaining construction: the final verb must be a motion verb and it must directly follow the preceding verb. In this construction, the vowel of the participle may be either long or short, with no noticeable difference in meaning:

- (6) *wā-ā* *nay-ā(ā)* *ma-wā-ā*
 3S-ERG eat-(L.)PART NEG-go-STAT
 'He ate and didn't leave.' OR 'He didn't eat and left' (i.e., 'he left without eating')

There is more to say about these constructions and the differences between these speakers (see Genetti 1986 for a fuller description), but this much is sufficient to make the following point: speakers may appear to have very similar speech on initial observation, but prove to have significant and systematic differences at a quite deep grammatical level.

To what should one attribute the differences between these speakers? One obvious approach, and the one I took in my original analysis, is to assume that the two Newārs speak different dialects. This analysis is in accordance with the expectation of interspeaker consistency. Kāthmāndu Newār society is highly complex and stratified. Members of the community are differentiated by caste, religion, occupation, gender,

socioeconomic and educational level, and neighborhood. The two speakers in question are from different castes and different parts of the city, such that an analysis of different dialects is certainly plausible. However, there are not other many obvious differences in the speech of these two consultants.

Another possible factor to which one might attribute this variation is the data elicitation technique. These data were collected in the artificial setting of a university office far removed from the speech community, out of context, and as translations of possible English sentences. So perhaps the difference in interpretation was due to one or both speakers losing track of their native intuitions in their role as informants.

A third possibility is that the variation is idiolectal, and that each speaker has independently constructed different grammatical systems of clause combining even though they had similar input as children. This analysis goes against the expectation of inter-speaker consistency, however, given my subsequent experience of working with significant idiolectal variation in Dolakhā Newār (discussed below), it has become clear to me that idiolectal variation is pervasive, and one cannot assume dialectal distinctions without clear independent evidence that the dialects exist.

3 Idiolectal variation in Dolakhā Newār

The expectation of inter-speaker consistency dominated my view of language until my second field trip to Nepal to study Dolakhā Newār. During my first field trip I had worked with one primary consultant named Kalpanā Shrestha, a young woman who had lived in Dolakhā until the age of twelve, then moved to Kāthmāndu with her family, where she was actively involved in the Dolakhā Newār community there and used the language regularly. Working with her, I collected vocabulary, conducted elicitation to determine the basic outlines of the grammar, and transcribed and translated a number of narratives produced by a variety of native speakers. When I returned to continue my work the following year, Kalpanā had taken a job and had little time to spare for work with me. She introduced me to Rama, her cousin two years older. With her, I also collected vocabulary, conducted elicitation and transcribed recorded texts, although I continued to work with Kalpanā when possible.

Kalpanā and Rama are first cousins. Their fathers are brothers, and, following the traditional Newār pattern, when they married they each brought their new wives into their shared house, and raised their children together. Thus, the two girls grew up in exactly the same linguistic community until Rama was seven and Kalpanā was five. At that point, Rama's father took a job with the government and was posted to other parts of Nepal. However, the family continued to consistently speak Dolakhā Newār in their home. Thus there are neither substantial regional or socio-demographic differences between the two women. The only possible difference between their linguistic backgrounds is that Rama

left the village at the age of seven at which point she interacted primarily with her parents for several years, and Kalpanā left the village at twelve. However, they both continued to be active members of the Dolakhā Newār community and continued to use the language consistently. It thus came as a surprise to me when I found variation in their speech at the phonological and morphosyntactic levels. I will present here a few, rather simple, examples of the variation in the speech of these two first cousins. There are many other areas in which their speech varies.

3.1 *Variation in phonology*

To begin with a simple example, Kalpanā pronounces the first person exclusive pronoun as /chiji/ whereas Rama pronounces it as /thiji/, although both women clearly have phonemic systems which differentiate between the aspirated alveolar stop /th/ and the aspirated alveopalatal affricate /ch/. This is a clear, if perhaps minor, violation of the expectation of inter-speaker consistency. When I pointed this difference out to the women one time when we were all together, they were both quite surprised and each laughingly insisted that her pronunciation was correct. In describing the pronominal system of the language then, the linguist is faced with a choice. Either choose one of the pronunciations as 'correct' and representing the pronoun, or list both pronunciations, thereby allowing for variation in the linguistic description. The problem with the first path is how to choose the 'correct' or 'basic' version. Perhaps one could do a survey of a representative sample of speakers, and choose the form that occurred most frequently, or one could choose /ch/ for systemic reasons as it corresponds to the second person pronoun *chi*, from which *chiji* is transparently formed (and for which there is no alternative *thi* to my knowledge). Or, one could choose *thi*, since it is more likely that this is historically prior, as the vowel /i/ together with the aspiration creates the ideal environment for palatalisation. But, if our linguistic descriptions are meant to be close and detailed portraits of the state of the language as we find it, then the most accurate description is the one that admits and highlights the variation. We may lose something in descriptive elegance, but we gain a more accurate and realistic portrait.

It is interesting to consider the *thiji/chiji* variation in light of the pronominal paradigms of which the forms are part. Although /th/ and /ch/ are clearly distinct phonemes and differentiate a number of nouns and verbs, they actually do not contrast forms within the narrow lexical class of personal pronouns. Assuming that speakers are aware when pronouns are being used in natural discourse—from their syntactic positioning, inflection, and discourse functions—they then do not have to maintain a strict phonemic distinction between similar phonemes. In this environment, the phonemic distinction can be relaxed, and variation easily tolerated without problems of intelligibility.

More extensive phonological variation between the two consultants is found in their vowel harmony systems. Dolakhā Newār has three verbal prefixes, the negative *ma-*, the

prohibitive *da-*, and the optative *tha-*. The vowels in all three prefixes are subject to harmony depending on the vowel of the stem. Rama has the simplest system of vowel harmony, as only the vowel /ā/ triggers harmony, for example *mā-y ā* ‘didn’t come’. Harmony is blocked when the stem has a glide in C2 position, for example, *ma-syāt* ‘didn’t kill’, *ma-mwāl* ‘didn’t search’. Kalpanā’s system is more complicated, as /ā/, /o/, and the sequence /wā/ all trigger harmony, thus, *dā-dāu* ‘don’t beat’, *mo-sou* ‘didn’t see’, *mwā-mwāl* ‘didn’t search’. A third consultant, a much younger cousin of the first two, has a more extensive system yet, with the feature [round] harmonising from /u/: *mo-pul* ‘didn’t pay’.

It is clear that all three speakers have formed distinct vowel harmony systems, but, as with the different pronunciations of the pronoun, none of them indicated awareness of any differences between their own speech and that of their cousins. Although the different vowels that appear in the prefixal allomorphs are distinct phonemes, they actually have no phonemic value in the prefixes themselves; these are the only three prefixes in the entire language and they are differentiated by the initial consonant. Thus the vowel carries no functional load and its precise quality is unimportant. As with the case of the pronoun, it thus appears that inter-speaker variation is especially tolerated in environments where it does not matter.

It is interesting to consider variation such as this in historical terms. Could this variation imply that vowel harmony in Dolakhā Newār is undergoing a change not yet completed, perhaps becoming more elaborate? If so, then we have an explanation for this counter-example to the expectation of inter-speaker consistency. It is simply a system in change, and over a matter of time inter-speaker consistency will once again be achieved. Of course this is a real possibility, and only time will tell, but there is no reason to assume that this is the case. It is just as likely that the reason for the synchronic variation is that the input that these speakers had as children was also variable, and that variation itself is a stable property of the language that persists over time.

3.2 Variation in morphosyntax

There are two nominalising suffixes in Dolakhā Newār, which are used in the formation of relative clauses, in complements of perception verbs, in complements of cognition verbs, in complements of *ju-en con-a* (roughly a mirative expression), in emphatic constructions, and in some types of questions. The analytical difficulty comes in determining under what conditions each of the two suffixes appear in these various environments. A full discussion is beyond the scope of the current paper (see Genetti 1994:154–170). Here I will begin by discussing the inflection of the verb in relative clauses. Note that since there are no simple functional terms which can be used to label these suffixes, I have resorted to calling them NR1 and NR2, for ‘nominaliser/relativiser’ 1 and 2 respectively.

When I first began working with Kalpanā on the distribution of these forms in relative clauses, I found her system to be quite straight-forward. NR1 was used in subject relative clauses (7), while NR2 was used in object relative clauses (8).

- (7) *chē=ku* *ye-u* *mi-pen*
 house=LOC come-NR1 person-PL
 'people who came to the house'
- (8) *jin* *khoy-a* *keṭi*
 1S.ERG see-NR2 girl
 'the girl whom I saw'

This distribution was borne out in text counts using texts from a number of speakers; out of one hundred and twenty subject and object relative clauses examined, there was only one counter-example to this pattern, and that from a speaker who was openly criticised by others in the room for mixing up the story as she told it. Given this clean, motivated pattern in my data, it was thus surprising to find that Rama's opinions on the possible distribution of the suffixes differed from Kalpanā's. For subject relatives, Rama preferred to use NR1 consistently, but she also allowed the possibility of NR2 in examples where the aspect of the clause was imperfective, as in (9).

- (9) *ām* *āmp* *kha-en* *coṅ-a / co-gu* *mucā*
 that mango pick-PART stay-NR2 / stay-NR1 child
 'the child picking mangos'

For object relatives, Rama again preferred NR2, but she said that NR1 was also possible in examples when the subject of the relative clause is third person, as in (10):

- (10) *āmun* *khoy-a / khoy-gu* *mi-pen*
 3S.ERG see-NR2 / see-NR1 person-PL
 'the people that he saw'

Kalpanā, however, did not accept either of these optional patterns, and insisted that NR1 was the only form possible for the example in (9), and that NR2 was the only form possible for (10).

My first inclination upon seeing this variation was to attribute it to the elicitation setting. One must admit it is a rather strange task to translate series of sentences such as 'I saw the man who gave Sita the money' and 'I saw the money that the man gave Sita', and maybe it would be easy for a consultant to get mixed up. In short, I assumed that Rama's seemingly unmotivated system was probably due to error on her part. However, her opinions remained consistent over time. I then began to elicit oblique relative clauses from both consultants, and found that while both 'preferred' NR2 consistently in these examples, they both admitted that it was possible to use NR1 with oblique relatives if the subject of the clause is third person, thus:

- (11) *ām* *misāmi* *ye-e / ye-u* *gāū*
 that woman come-NR2 / come-NR1 village
 'the village that the woman came from'

Text counts of oblique relative clauses with third-person subjects did indeed show such variation; while twenty-two examples had NR2 (judged 'preferred' by both consultants), five examples had NR1. Looking back to Rama's earlier judgments on object relatives, I noticed that the pattern she had given me for the inflection of the verb in object relative clauses was the same pattern that she and Kalpanā had both given me for the inflection of the verb in oblique relatives, and for which there was evidence in the texts. Thus I began to consider the possibility that, rather than being confused about the inflection of the verb in object relatives, Rama's judgments could result from a more finely tuned awareness of variation in the use of these forms.

In examining the distribution of NR1 and NR2 in other environments, I found that different factors condition their appearance in different environments: With complements of perception verbs, verbal transitivity is important; with complements of cognition verbs, perfectivity is the conditioning feature; in questions and emphatic constructions, person of the subject is relevant; in complements of *juen cona*, both transitivity and perfectivity are the factors speakers attend to. What emerges is a complex set of variables which underlies the distribution of the two paradigmatic suffixes, but no clearly motivated pattern. While the distinction between the suffixes can probably be traced to an old historical distinction based on transitivity (see Genetti 1994:169–171), there is no evidence for this synchronically. The attested variation appears, then, to be due to a decayed system which lacks a coherent functional motivation for the two suffixes, and whose use has thus become idiosyncratic and variable. Faced with this type of situation, speakers may choose to reanalyze and regularise the system. This seems to be happening in the case of relative clauses, the most frequent environment where the suffixes occur, and which appear for some speakers to have moved into a motivated system based on grammatical relations.

4 Implications of the studies

In the discussion of the case studies, I have made a number of points about the nature of linguistic variation, which may be summarised as follows:

- Variation may persist across generations
- Variation may result from competing motivations, which are resolved differently by different speakers, or result in register-based variation
- Speakers may appear on the surface to have very similar linguistic systems, but differences may emerge at quite deep levels of the grammar
- Significant idiolectal variation may exist in the phonology, morphology and syntax, even among speakers with identical linguistic backgrounds
- One should not attribute variation to dialect without independently proving the existence of the dialects; in the absence of this evidence, the variation may be idiolectal as opposed to dialectal

- Idiolectal variation may be more likely to arise in cases where it makes no difference, that is, in cases where there is no functional loss resulting from the variation
- Variation may reflect systems in the process of historical change, but doesn't necessarily have to.

It is clear that one may draw many deep insights into the nature of language by exploring linguistic variation, and that incorporating variation into linguistic description allows for a richer understanding of language as well as a more accurate portrayal. Our tradition in linguistic theory is to expect inter-speaker consistency, and to assume that we can produce a single accurate statement about linguistic systems and subsystems. Variation is often attributed to dialect or register only, and so 'accounted for', often with little in-depth exploration or justification. With this approach we risk obscuring the richness and diversity of language in our search for generalisations. We are fortunate to have the work of Jim Matisoff as a model of an alternative approach, which inspires us not just to work with unkempt variation, but to revel in it.

References

- Genetti, Carol, 1986, Scope of negation in Newari clause chains. Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics, Ohio State University, 11–14 September 1986.
- 1994, *A descriptive and historical account of the Dolakha Newari dialect* (Monumenta Serindica). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- 1999, Variation in agreement in the Nepali finite verb. In Warren Glover and Yogendra Yadava eds *Topics in Nepalese linguistics*, 542–555. Kathmandu: Royal Nepal Academy. (First published 1993, *South Asian Language Review* 3/2:90–104.)
- Thomason, Sarah Grey and Terrence Kaufman, 1988, *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Turner, Ralph Lily, 1921, Specimens of Nepali. *The Indian Antiquary* 50:84–92.
- 1922, Further specimens of Nepali. *The Indian Antiquary* 51:41–67.

